

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan media *online* yang begitu pesat, memunculkan banyak media mulai dari media berita sampai media sosial. Media sosial saja sudah begitu banyak, dari *Facebook*, *Twitter*, *Path*, *Instagram*, *Google+*, *Tumblr*, *Linkedin* dan masih banyak lagi[1]. Pengguna aktif harian *twitter* di indonesia saat ini cukup banyak, saat ini diberbagai belahan dunia *Twitter* diklaim menjadi salah satu media *online* yang pengguna aktif harian yang paling banyak berdasarkan laporan finansial *twitter* pada kuartal ke-3 tahun 2019[2]. Melihat perkembangan *twitter* tersebut maka *twitter* menjadi salah satu media yang dapat digunakan untuk melakukan analisis sentimen terhadap berbagai topik. Penelitian ini melakukan suatu analisis sentimen terhadap bahasan yang saat ini sering menjadi *trending topic* di twitter yaitu “CoronaVirus Disease-2019 (covid19)”.

COVID-19 adalah penyakit menular yang disebabkan oleh jenis coronavirus yang baru ditemukan. Ini merupakan virus baru dan penyakit yang tidak dikenal sebelum terjadi wabah di Wuhan, Tiongkok, bulan Desember 2019[3]. Penyebaran *virus* ini juga langsung dibicarakan oleh banyak kalangan masyarakat twitter, saat ini *virus corona* tengah menjadi perhatian dunia internasional. Banyaknya jumlah angka korban dan cepatnya penularan virus membuat masyarakat khawatir dan muncul berbagai opini tentang *virus corona*, Opini inilah yang kemudian di analisa untuk diketahui polaritasnya dengan analisis sentimen.

Penelitian menggunakan machine learning untuk mengklasifikasikan movie reviews(Pang, 2002). Penelitian ini pernah dilakukan dengan cara mengklasifikasikan sentimen terhadap review film dan menentukan apakah review film tersebut memiliki sentimen positif ataupun negatif. Fitur yang berbeda dari review tersebut diekstrak dan digunakan algoritma *machine learning* Naïve Bayes dan *Support Vector Machine* (SVM) untuk menghasilkan model klasifikasi. Mereka memperoleh hasil akurasi antara 78,7% ketika menggunakan Naïve Bayes dan nilai akurasi yang diperoleh ketika menggunakan SVM adalah 72,8%.

Berdasarkan latar belakang yang telah dipaparkan, penulis tertarik untuk melakukan penelitian dengan judul “**Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)**”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan diatas maka dapat dibuat rumusan masalah :

1. Bagaimana metode *Logistic Regression* Dan *Support Vector Machine(SVM)* mampu secara otomatis melakukan klasifikasi *Twitter* dengan baik?
2. Berapa akurasi yang diperoleh dari metode *Logistic Regression* Dan *Support Vector Machine(SVM)* untuk melakukan klasifikasi *Twitter*?

1.3 Batasan Masalah

Agar penelitian yang dilakukan menjadi lebih spesifik sesuai perumusan masalah tersebut maka batasan yang ditentukan adalah sebagai berikut :

1. *Tweet* yang diambil dan dianalisis hanya *tweet* berbahasa Inggris Dari *Keaggle*.
2. Metode yang dipakai adalah *Logistic Regression* Dan *Support Vector Machine(SVM)*.
3. Sentimen *Twitter* hanya sentimen yang berhubungan dengan *Corona Virus Disease-2019 (Covid19)*.
4. *Tweet* hanya dikelompokkan kedalam 5 sentimen *Extremely positive* , *positive*, *neutral*, *negative*, *Extremely negative*.

1.4 Tujuan Penelitian

Tujuan diadakannya penelitian yaitu untuk :

1. Mengetahui metode *Logistic Regression* Dan *Support Vector Machine(SVM)* mampu secara otomatis melakukan klasifikasi *tweet* pada *Twitter* dengan baik.
2. Mengetahui akurasi yang diperoleh dari metode *Logistic Regression* Dan *Support Vector Machine(SVM)* untuk melakukan klasifikasi *Twitter*.

1.5 Manfaat Penelitian

Manfaat yang dapat diberikan penelitian ini adalah :

1. Membantu menganalisis sentimen *tweet* pada *Twitter* menggunakan metode *Logistic Regression* Dan *Support Vector Machine(SVM)*.
2. Menjadi referensi bagi peneliti mendatang bila akan melakukan penelitian dengan tema terkait.